Listing of Claims:

1. (Previously Presented) A method for generating a language component vocabulary VC for a speech recognition system having a language vocabulary V of a plurality of word forms, the method comprising the steps of:

partitioning the language vocabulary V into subset of word forms based on frequencies of occurrence of the respective word forms;

in at least one of said subsets, splitting word forms having frequencies less than a threshold to thereby generate word form components; and

generating a language component vocabulary VC comprising word forms and word form components.

2. (Original) The method of claim 1, wherein the frequencies of the word forms are estimated from a given textual corpus.

3. (Original) The method of claim 1, wherein said partitioning step includes the sub-step of numerating the plurality of word forms in the language vocabulary V in descending order based on the frequencies associated with each of the plurality of word forms.

4. (Original) The method of claim 1, wherein said partitioning step partitions the language vocabulary V into at least two subsets S1 and S2, and said splitting step splits the word forms of subset S2 into 2-tuple components including stems and endings, but does not split the word forms of subset S1.

5. (Original) The method of claim 4, wherein said partitioning step further partitions the language vocabulary V into a third subset S3, with word forms therein being split in said splitting step into 3-tuple components including prefixes, stems and endings.

6. (Original) The method of claim 1, wherein said splitting is performed subject to a constraint in which a word that contains a given string of letters is prevented from being split within the string if the string of letters corresponds to one phoneme.

7. (Original) The method of claim 1, wherein said splitting is performed using a fixed vocabulary and a fixed list of allowable endings, with each word from the fixed vocabulary being split into at least a stem and an ending that is an element of the fixed set of endings, so as to substantially minimize the total number of all stems that are required to split every word in the fixed vocabulary.

8. (Original) The method of claim 7, wherein the fixed set of allowable endings includes an empty ending.

9. (Original) The method of claim 1, further comprising generating and storing a word form to corresponding word form components table.

10. (Original) The method of claim 9, further comprising the step of labeling each of the word form components stored in said table to distinguish between stems, prefixes and endings.

11. (Original) The method of claim 1, further comprising the steps of:

generating a map of said word forms to said word form components, said map further including each of a plurality of non-split words as being associated with itself;

filtering a textual corpus using the map to generate a textual component corpus containing the non-split word forms and the word form components of the map;

accumulating the word form components and the non-split word forms generated by said filtering step in an n-gram language model; and

determining counts of n-tuple sets of word form components and word forms to estimate n-gram probabilities for the n-gram language model.

12. (Original) The method of claim 11, wherein said filtering step maps every word in the corpus into a n-tuple word form component.

13-41. (Canceled)

42. (Original) A method for splitting words in a language vocabulary V in an automatic speech recognition system to provide vocabulary compression, wherein the vocabulary V has a fixed size, the method comprising the steps of:

(a) providing a fixed set of allowable endings, including an empty ending;

(b) providing a fixed set of constraints for splitting words into stems;

(c) initializing a split map of words and the corresponding stems and endings by setting a variable t to a predetermined value, and selecting a first word from the fixed vocabulary;

(d) randomly splitting the first word to generate an ending from the fixed list of allowable endings and a stem;

(e) defining and storing a stem set containing the stem generated at said splitting step (d) and a word set containing the first word;

(f) determining whether t is less than the size of the vocabulary V;

(g) obtaining a new word from the vocabulary V, when t is less than the size of the vocabulary V;

(h) determining possible splits for the new word to generate stems and endings therefrom, using the fixed set of allowable endings and the fixed set of constraints;

(i) determining whether there is a split for the new word that generates a previously stored stem of the stem set;

(j) splitting the current word into the previously stored stem and an ending of the set of allowable endings, when there is a split for the new word that generates the previously stored stem of the stem set;

(k) determining whether another previously stored stem in the stem set can be replaced by a new stem generated at step (h), when there is no split for the current word that generates the previously stored stem of the stem set;

(l) redefining the stem set and the split map to include the new stem generated at step (h) in place of the other previously stored stem, when the other previously stored stem can be replaced by the new stem, when the other previously stored stem can be replaced by the new stem generated at step (h);

(m) redefining the stem set to include any new stem into which the current word may be split and extending the split map to include the current word by splitting the new word into the

new stem, when the other previously stored stem in the stem set cannot be replaced by the new stem generated at step (h); and

(n) incrementing t and returning to step (f) if t is less than the size of the vocabulary V.

43. (Original) The method of claim 42, further comprising the step of terminating the method if t is not less than the size of the fixed vocabulary.

44. (Original) The method of claim 42, wherein said determining step (k) comprises the step of determining whether other words stored in the word set during previous iterations will remain split after such substitution.

45. (Original) The method of claim 42, wherein the vocabulary is sorted such that the words in the language vocabulary V are numerated in descending order based on frequencies associated with each of the words.

46. (Original) The method of claim 42, wherein step (j) further comprises the step of extending the split map to the new word.

47. (Original) The method of claim 42, wherein step (i) generates all possible splits for the new word.

48. (Previously Presented) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for generating a language component vocabulary VC for a speech recognition system having a language vocabulary V of a plurality of word forms, the method steps comprising:

partitioning the language vocabulary V into subset of word forms based on frequencies of occurrence of the respective word forms;

in at least one of said subsets, splitting word forms having frequencies less than a threshold to thereby generate word form components; and

generating a language component vocabulary VC comprising word forms and word form components.

49. (Previously Presented) The program storage device of claim 48, wherein the frequencies of the word forms are estimated from a given textual corpus.

50. (Previously Presented) The program storage device of claim 48, wherein the instructions for partitioning comprise instructions for numerating the plurality of word forms in the language vocabulary V in descending order based on the frequencies associated with each of the plurality of word forms.

51. (Previously Presented) The program storage device of claim 48, wherein the instructions for partitioning comprise instructions for partitioning the language vocabulary V into at least two subsets S1 and S2, and wherein the instructions for splitting comprise instructions for

splitting the word forms of subset S2 into 2-tuple components including stems and endings, but not splitting the word forms of subset S1.

52. (Previously Presented) The program storage device of claim 51, wherein the instructions for partitioning further comprise instructions for partitioning the language vocabulary V into a third subset S3, with word forms therein being split in said splitting step into 3-tuple components including prefixes, stems and endings.

53. (Previously Presented) The program storage device of claim 48, wherein the instructions for splitting comprise instructions for performing the splitting subject to a constraint in which a word that contains a given string of letters is prevented from being split within the string if the string of letters corresponds to one phoneme.

54. (Previously Presented) The program storage device of claim 48, wherein the instructions for splitting comprises instruction for performing the splitting using a fixed vocabulary and a fixed list of allowable endings, with each word from the fixed vocabulary being split into at least a stem and an ending that is an element of the fixed set of endings, so as to substantially minimize the total number of all stems that are required to split every word in the fixed vocabulary.

55. (Previously Presented) The program storage device of claim 54, wherein the fixed set of allowable endings includes an empty ending.

56. (Previously Presented) The program storage device of claim 48, further comprising instructions for generating and storing a word form to corresponding word form components table.

57. (Previously Presented) The program storage device of claim 56, further comprising instructions for labeling each of the word form components stored in said table to distinguish between stems, prefixes and endings.

58. (Previously Presented) The program storage device of claim 48, further comprising instructions for:

generating a map of said word forms to said word form components, said map further including each of a plurality of non-split words as being associated with itself;

filtering a textual corpus using the map to generate a textual component corpus containing the non-split word forms and the word form components of the map;

accumulating the word form components and the non-split word forms generated by said filtering step in an n-gram language model; and

determining counts of n-tuple sets of word form components and word forms to estimate n-gram probabilities for the n-gram language model.

59. (Previously Presented) The program storage device of claim 58, wherein the instructions for filtering comprise instructions for mapping every word in the corpus into a n-tuple word form component.

60. (Previously Presented) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for splitting words in a language vocabulary V in an automatic speech recognition system to provide vocabulary compression, wherein the vocabulary V has a fixed size, the method comprising the steps of:

(a) providing a fixed set of allowable endings, including an empty ending;

(b) providing a fixed set of constraints for splitting words into stems;

(c) initializing a split map of words and the corresponding stems and endings by setting a variable t to a predetermined value, and selecting a first word from the fixed vocabulary;

(d) randomly splitting the first word to generate an ending from the fixed list of allowable endings and a stem;

(e) defining and storing a stem set containing the stem generated at said splitting step (d) and a word set containing the first word;

(f) determining whether t is less than the size of the vocabulary V;

(g) obtaining a new word from the vocabulary V, when t is less than the size of the vocabulary V;

(h) determining possible splits for the new word to generate stems and endings therefrom, using the fixed set of allowable endings and the fixed set of constraints;

(i) determining whether there is a split for the new word that generates a previously stored stem of the stem set;

(j) splitting the current word into the previously stored stem and an ending of the set of allowable endings, when there is a split for the new word that generates the previously stored stem of the stem set;

(k) determining whether another previously stored stem in the stem set can be replaced by a new stem generated at step (h), when there is no split for the current word that generates the previously stored stem of the stem set;

(l) redefining the stem set and the split map to include the new stem generated at step (h) in place of the other previously stored stem, when the other previously stored stem can be replaced by the new stem, when the other previously stored stem can be replaced by the new stem generated at step (h);

(m) redefining the stem set to include any new stem into which the current word may be split and extending the split map to include the current word by splitting the new word into the new stem, when the other previously stored stem in the stem set cannot be replaced by the new stem generated at step (h); and

(n) incrementing t and returning to step (f) if t is less than the size of the vocabulary V.